Cell, Vol. 82, 209–220, July 28, 1995, Copyright © 1995 by Cell Press

Structure of the Bacteriophage Mu Transposase Core: A Common Structural Motif for DNA Transposition and Retroviral Integration

Phoebe Rice and Kiyoshi Mizuuchi
Laboratory of Molecular Biology
National Institute of Diabetes
and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20892-0540

Summary

The crystal structure of the core domain of bacteriophage Mu transposase, MuA, has been determined at 2.4 Å resolution. The first of two subdomains contains the active site and, despite very limited sequence homology, exhibits a striking similarity to the core domain of HIV-1 integrase, which carries out a similar set of biochemical reactions. It also exhibits more limited similarity to other nucleases, RNase H and RuvC. The second, a β barrel, connects to the first subdomain through several contacts. Three independent determinations of the monomer structure from two crystal forms all show the active site held in a similar, apparently inactive configuration. The enzymatic activity of MuA is known to be activated by formation of a DNA-bound tetramer of the protein. We propose that the connections between the two subdomains may be involved in the cross-talk between the active site and the other domains of the transposase that controls the activity of the protein.

Introduction

Bacteriophage Mu, which multiplies by replicative transposition, is one of the best characterized of the mobile genetic elements (for reviews see Haniford and Chaconas, 1992; Mizuuchi, 1992). The transposase responsible for this reaction, MuA, is a member of a large family of transposase and integrase enzymes. The reactions catalyzed by these enzymes are important, among other things, in the transmission of drug resistance among bacteria, in the integration of retroviruses, and as tools for genetic engineering and gene therapy.

The Mu transposition reaction proceeds through a series of protein–DNA complexes called Mu transpososomes, at the core of which are the two ends of the Mu genome, synapsed by a tetramer of the MuA protein. MuA first introduces a site-specific nick at each end of the Mu genome and then catalyzes the attack of the newly freed 3′ OHs on a target DNA molecule, resulting in a branched DNA structure, which can be assembled into a replication fork by recruiting host enzymes (Figure 1; Craigie and Mizuuchi, 1987; Surette et al., 1987; Kruklitis and Nakai, 1994). Aberrant transposition reactions could be fatal to the bacteriophage, and under physiological conditions, assembly of the proper complex is tightly controlled. Some of these control factors can be circumvented in vitro, but chemical activity of the protein is still tightly correlated with formation of tetramers that synapse two Mu end DNA segments (Mizuuchi et al., 1992; Baker and Mizuuchi, 1992; Namgoong et al., 1994; H. Savilahti, P. R., and K. M., unpublished data).

MuA is 75 kDa in size, is monomeric in the absence of DNA (Baker and Mizuuchi, 1992; K. M., unpublished data), and can be divided into several domains. The first 76 residues bind to an internal activating sequence (IAS, an enhancer-like sequence element required for efficient assembly of the Mu transpososome) that is transiently bound during complex assembly but is not stably bound to the final transpososome (Leung et al., 1989; Surette and Chaconas, 1992; Mizuuchi and Mizuuchi, 1989). The structure of this domain has recently been shown by nuclear magnetic resonance (NMR) to be a winged helix-turn-helix motif (Clubb et al., 1994). The next domain, residues 77–247, is responsible for sequence-specific binding to the Mu genome ends (Nakayama et al., 1987; Leung et al., 1989). Residues 248–574 form the core of the enzyme that contains the active site and nonspecific DNA-binding activity (Baker and Luo, 1994; Nakayama et al., 1987). The C-terminal domain, residues 575–663, contains a region that is important in tetramer formation, 575–605 (Desmet et al., 1989; Betermier et al., 1989; Baker et al., 1993), and also a region that is responsible for interaction with MuB protein, an allosteric activator of MuA that also binds to target DNA and facilitates target immunity (Wu and Chaconas, 1994; Baker et al., 1991; Surette and Chaconas, 1991; Surette et al., 1991; Adzuma and Mizuuchi, 1988). MuA protein truncated at residue 574 will not form tetramers on its own, but will participate in the reaction if mixed with intact protein (Baker et al., 1993; T. A. Baker, personal communication). Thus, this version of the enzyme is active but partially defective in oligomerization.

In the MuA tetramer, each monomer appears to be responsible for a separate step of the reaction: two monomers nick the two Mu genome ends, and the other two catalyze the transfer of the ends into the target DNA (Baker et al., 1993, 1994). Both reactions require divalent metal ions, and for the strand transfer reaction, it has been demonstrated that the reaction proceeds with inversion of chirality at the phosphate, implying that the reaction proceeds by direct attack of the 3′ OH on the target DNA rather than through a protein–DNA covalent intermediate (Mizuuchi and Adzuma, 1991). For the closely related HIV integrase reaction, inversion of chirality has been demonstrated at both the viral DNA end cleavage and strand transfer steps (Engelman et al., 1991).

Mutational analysis has identified three acidic residues that are essential for catalytic activity and that have been implicated in metal binding: D269, D336, and E392 (Baker and Luo, 1994; Kim et al., 1995; E. Krementsova and T. A. Baker, personal communication). These residues lie in the core domain and correspond to the D–D–35–E motif that is conserved among many transposases and retroviral integrases (Kulkosky et al., 1992; Radstrom et al., 1994;

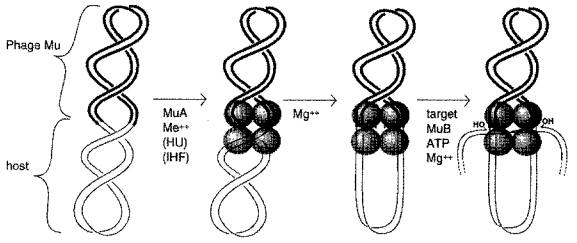

Figure 1. Diagram of Transposition Pathway

MuA binds specifically to the ends of the bacteriophage Mu genome and forms a tetramer that synapses the two ends of the genome. Under physiological conditions, divalent metal ion, supercoiling, and the host proteins HU and IHF are needed for efficient complex formation. An "internal activating sequence" (not shown) found within the Mu genome is transiently bound during assembly. In the presence of Mg^{2+}, specific nicks are introduced at the Mu–host junction. The target DNA is "delivered" by the MuB protein. MuB is a nonspecific DNA-binding ATPase that mediates target immunity and also activates MuA for strand transfer. The 3′ OH groups at the ends of the Mu genome attack the target DNA to generate the strand transfer product, which is resolved by host proteins.

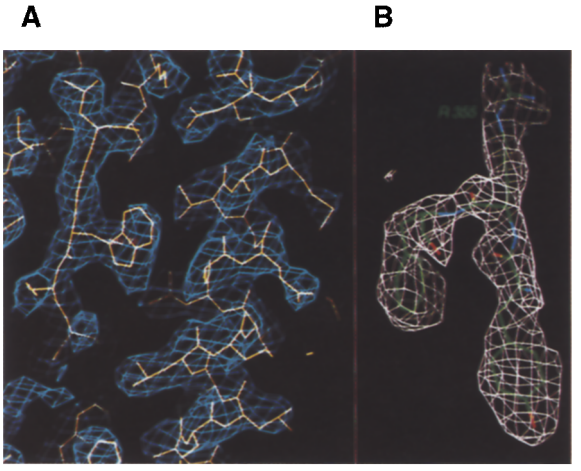

Figure 2. Electron Density

(A) Experimental electron density map for the C2 crystal form. Refined coordinates are shown superimposed on the averaged experimental electron density map calculated at 3.1 Å resolution. Portions of β strand 3 and helix E are shown. See Experimental Procedures for details.

(B) Simulated annealing omit map from the $C222_1$ crystal form. Residues 354–356 were deleted from the model, and all residues within 3 Å of them were restrained to their original positions. The $F_{obs} - F_{calc}$ map, after simulated annealing, is shown contoured at 2.4 σ. Although R355 lies in a disallowed region of the Ramachandran plot, clear density can be seen for the peptide backbone.

Doak et al., 1994), although the spacer between the second D and the E is 55 rather than 35 amino acids in MuA.

We report here the crystal structure of the core domain of MuA, residues 248–574. Two different crystal forms of the protein have been solved, one at 2.8 Å resolution and one at 2.4 Å. Our structure can be divided into two subdomains, the first of which is strikingly similar to the structure of the core domain of HIV-1 integrase (Dyda et al., 1994). While some structural similarity, at least in the active site region, was expected from the mechanistic similarities between the two enzymes, the degree of similarity seen is much greater than could have been anticipated from the very limited sequence homology. Our structure and the recently published integrase structure reveal that these enzymes are not only related to one another, but also to a larger family of polynucleotidyl transferases, including RNase H and the Holliday junction–resolving enzyme RuvC (Katayanagi et al., 1990, 1992; Yang et al., 1990; Ariyoshi et al., 1994; Davies et al., 1991).

Results

Structure Solution

The structure of the core domain of MuA has been solved in two different crystal forms. The monoclinic crystal form, which contains two monomers in the asymmetric unit, was solved by multiple isomorphous replacement, using two heavy metal derivatives and the anomalous scattering component from one of them (see Experimental Procedures for details). Initial phases were calculated to 3.2 Å and were improved and extended to 3.1 Å by solvent flattening and averaging about the noncrystallographic twofold axis. (Figure 2A). It has now been refined to an R factor of 21.4% at 2.8 Å resolution, with good geometry (Table 1).

The orthorhombic crystal form contains one monomer in the asymmetric unit and was solved by molecular replacement, using monomer 1 from the first crystal form as a search model. This structure has been refined to an R factor of 19.6% at 2.4 Å, also with good geometry.

Three independent determinations of the monomer structure are provided by these two crystal forms. In all three, 9 or 10 residues at the N-terminus and 14 residues at the C-terminus are not visible in the electron density map and are therefore not included in the model. In the orthorhombic crystal form, one turn (418–425) is also disordered. In the Ramachandran plot for the higher resolution structure, 89.7% of residues fall into the most favored regions (as defined by PROCHECK, Laskowski et al., 1993). Only one nonglycine residue, R355, lies in a disallowed region in the Ramachandran plot. This residue is found in the same conformation in all three models, and electron density for it is clearly visible in a simulated annealing omit map calculated over this region, including well-defined carbonyl oxygens, confirming that the backbone conformation in our model is correct in this region (Figure 2B). Interestingly, the peptide bond immediately N-terminal to this residue appears to be a predominant chymotrypsin cleavage site in the Mu transpososome (T. A. Baker, personal communication; see Discussion).

Overall Structure and the Active Site

The MuA core can be divided into two subdomains: residues 258–490 form a mixed α/β domain, which contains the enzyme active site, and residues 491–560 form a smaller C-terminal β barrel (Figure 3). The N-terminal subdomain contains a central five-stranded mixed parallel and antiparallel β sheet, with helices on either side. Its structure bears a striking resemblance to that of several other enzymes that also catalyze the nicking/transfer of nucleic

Table 1. Refinement Statistics

	Form 1	Form 2
Space group	C2	$C222_1$
Monomers per asymmetric unit	2	1
Unit cell	a = 172.3 Å b = 52.2 Å c = 101.6 Å β = 115.5°	a = 55.8 Å b = 81.1 Å c = 161.3 Å
Resolution range	8–2.8 Å	8–2.4 Å
Completness of data[a]		
Overall	93.0%	87.0%
Outermost shell	84.7% (2.8–2.92 Å)	65.7% (2.4–2.51 Å)
R[b]	21.4%	19.6%
R_{free}[c]	27.6%	26.0%
Rms deviation from ideal		
Bond lengths	0.008 Å	0.011 Å
Bond angles	1.32°	1.49°

[a] Data with $F < 1\sigma_F$ were excluded from the refinement.
[b] $R = \Sigma|F_{calc} - F_{obs}|/\Sigma F_{obs}$ for all reflections used in the refinement.
[c] R_{free} = the same as R, but for the 5% of the reflections that was not used in the refinement process.

acid strands: HIV-1 integrase, RNase H, and RuvC (see below).

Two of the residues known to be essential for catalysis, D269 and D336 (Baker and Luo, 1994; E. Krementsova and T. A. Baker, personal communication), are found on adjacent strands of the β sheet, and the third, E392, is on a loop that passes across the front of the sheet. E392 lies too far from the two aspartic acids to be involved in cooperative binding to a divalent metal ion. We suspect that the fragment as we see it is in an inactive conformation, which may reflect the fact that MuA is not normally active until it has formed a tetramer bound to the proper DNA substrates. We imagine that the loop containing E392 rearranges to bring the three acidic residues in close proximity upon activation of the protein (see Discussion for details). The position of E392 in our model is not simply the effect of crystal packing contacts on an otherwise completely flexible or disordered loop, since it lies in very similar positions in both crystal forms and since in either form there are no packing contacts in the immediate vicinity. The face of the β sheet just above the active site (as shown in Figures 3B and 3C) is rather polar for its partially buried location and includes a network of hydrogen bonds involving residues Q293 (on strand 2) and N276 (on strand 1) and several water molecules.

The helix following E392 (αD) is quite unusual in its structure. It begins with a short stretch of 3_{10} helix (in which the hydrogen bonds extend from residue i to residue i+3), then forms one standard α helix hydrogen bond (i to i+4), and ends with a π helix–type hydrogen bond (i to i+5). Water molecules form hydrogen bonds to the peptide backbone at the transition points between hydrogen bonding patterns, where some of the donor and acceptor groups would otherwise be left unsatisfied. A simulated annealing omit map calculated over this region confirmed our interpretation (data not shown).

D294 is conserved in many related enzymes, and the substitution of N for D at this residue destroys all but a trace of the catalytic activity (Baker and Luo, 1994; Kim et al., 1995). This residue is not part of the active site, but rather is part of the tight turn between β strands 2 and 3. It forms hydrogen bonds both to a backbone amide in the turn and to the side chain of T297, which is generally conserved as S or T in related enzymes.

The C-terminal subdomain is an antiparallel six-stranded β barrel. The topology of this barrel, a Greek key followed by a hairpin, is the same as that of the two β barrels of chymotrypsin (Matthews et al., 1967), except that the MuA barrel is "inside out" relative to chymotrypsin: the face of the β sheet that forms the inside of the MuA barrel forms the outside of the chymotrypsin barrels.

Interactions between Subdomains

The interface between the two subdomains is not large and includes a number of partially buried water molecules. Most of the direct contacts of the first subdomain with the second are made by helix G and by the loop between helices B and C. The tip of this loop is structurally an integral part of the second subdomain, as three of its residues, Y354, F356, and V358, form part of the hydrophobic core of the second subdomain and are separated by partially buried water molecules from the rest of the first subdomain (Figure 4A). Furthermore, there are several main chain hydrogen bonds between this loop and β strands 6 and 7 of the second subdomain. The extended loop between β strands 1 and 2 also makes some, but considerably less, contact with the second subdomain. This loop also has large hydrophobic residues at its tip (W279 and F280), but they are more closely associated with the first subdomain than the second. Helix A and β strand 3 make largely water-mediated contacts with the second subdomain. In fact, a line of water molecules can be seen between these two secondary structure elements and the second subdomain that runs through the entire interface. The only direct contact from strand 3 to the barrel is a bidentate salt bridge from R304 to D536 (Figure 4B). Mutation of D536 to N has been shown to reduce the activity of the protein to 68% of wild type in an in vivo assay (Kim et al., 1995).

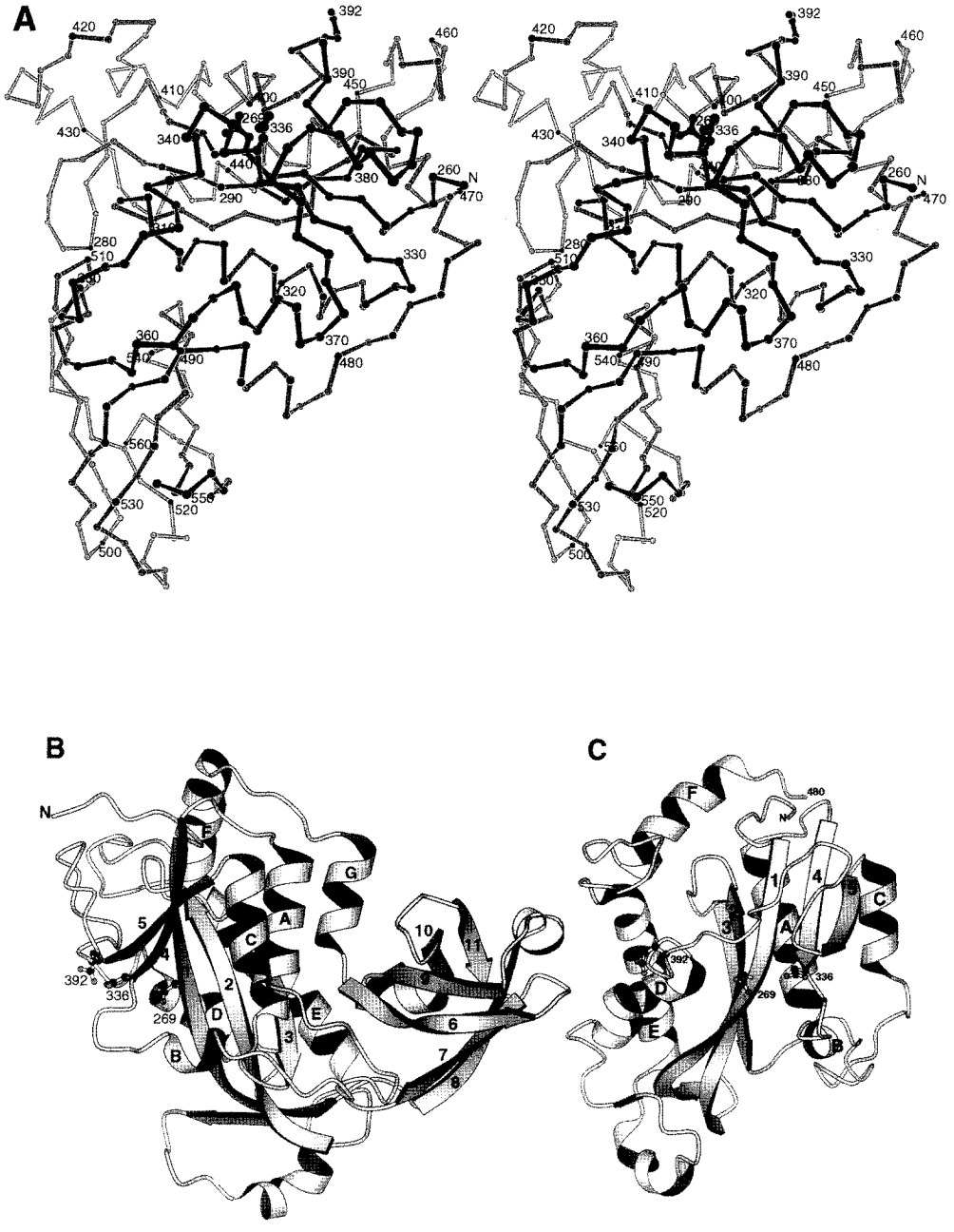

Figure 3. Overall Structure

(A) Stereo Cα drawing of the MuA core. Every tenth α carbon is highlighted in black.

(B) Ribbon drawing of the MuA core. Numbers given to the helices and strands are consistent with those assigned to the homologous HIV integrase core, and the three active site carboxylate residues are shown in black.

(C) Orthogonal view of the MuA core. View is approximately 90° away from that shown in (B). The second subdomain, which lies behind the first in this view, is omitted for clarity.

Comparison of the Three Independent Models

The two models from the C2 crystal form and the single model from the $C222_1$ crystal can be superimposed (as single rigid bodies) on one another with a root-mean-square (rms) deviation from the average position of 0.41 Å for all backbone atoms in common and 0.65 Å for all atoms in common. The major difference among the models is the relative orientation of the two subdomains. This differs by about 4° between the $C222_1$ model and monomer 2 of the C2 crystal (monomer 1 is between these two extremes). That this domain interface should be flexible is not surprising given its small size and rather hydrated nature.

Crystal Packing

In both crystal forms, each monomer makes two dyad-related interfaces. Each monomer interacts with its neighbors via the same two general regions of molecular sur-

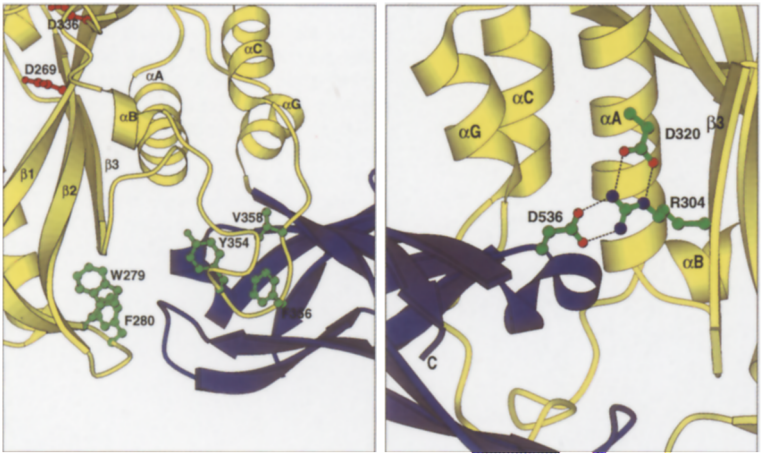

Figure 4. Interactions between Subdomains

(A) Closeup view of the domain interface. View is similar to that in Figure 3B, but rotated upward around a horizontal axis. The first subdomain is colored yellow and the second blue. The three hydrophobic residues at the tip of the loop between helices B and C (shown in green) form part of the hydrophobic core of the second subdomain. Two large hydrophobic residues at the tip of the loop between strands 1 and 2 (also in green) form a small part of the interaction surface between the two subdomains, but are more closely associated with the core of subdomain 1.

(B) View is approximately 180° from that shown in Figure 3B. R304 and D536 form a bidentate salt bridge between the subdomains.

face, but the specific orientations of the partners in each dimer vary. The largest interaction surface is found between monomers 1 and 2 of the C2 crystal form: using a 1.4 Å probe, this interface buries a total of 2060 Å^2 of previously solvent-accessible surface (Figure 5A). However, this figure may be slightly misleading, as the peripheral parts of the interface are not tightly packed and contain a number of buried water molecules. The same general interaction is repeated in the C222$_1$ crystal, but the two monomers have shifted and pulled away from one another such that only 890 Å^2 are buried (Figure 5B). In both cases, an important part of the interaction appears to be the burying of two leucines (368 and 369) found on the surface of helix C.

The other dyad-related interface is formed by portions of helices D and E and the turn between them. The exact contacts at this interface vary widely for all three monomers, resulting in 975 Å^2 of surface area buried for the dimer formed by monomer 2 of the C2 crystal with a symmetry-related monomer 2, 548 Å^2 for the dimer formed by two monomer 1s, and 1010 Å^2 for the C222$_1$ version.

Thus, in both cases, only the general region of protein surface used to make the interface, and not the particular contacts, is conserved.

Structural Similarity with Other Proteins

MuA is a member of a family of polynucleotidyl transferase enzymes that share a structurally related core. To date, the family includes HIV-1 integrase, RuvC, and RNase H (Dyda et al., 1994; Yang and Steitz, 1995). The central five-stranded β sheet and helices equivalent to helices A and C of MuA can be found, with the same topology, in all members of the family. All but RuvC also share helices B and D. Integrase and MuA are the most closely related members of the family: all the secondary structure elements found in the integrase core are also found in MuA (Figure 6). All of these enzymes are considered to use three (sometimes four) carboxyl groups to position one or two catalytic divalent metal ions. In all cases, two of these residues (usually aspartate) are found near the C-terminus of β strands 1 and 4 at a point where the strands split away from one another, forming a small cleft. The position of the third carboxyl, usually a glutamate, is more variable. In three of the four, it is found on the front side of the β sheet (as oriented in Figure 6A): for MuA and HIV integrase, just before helix D, and for RuvC on a helix that lies in the same general location as helix D but runs in the opposite direction. In RNase H, it is found on helix A, on the back side of the β sheet. The N-terminal part of helix D and the loop just before it in RNase H contain residues that, while not essential for catalysis, have considerable effects on K_{cat} and K_m when mutated (Haruki et al., 1994; Oda et al., 1993b; Kanaya et al., 1990).

These enzymes show a somewhat more distant relationship to both the kinase family and the 3′–5′ exonuclease domain of Klenow fragment (Yang, 1991; Artymiuk et al., 1993). These enzymes have five β strands and helices equivalent to helices A and C of MuA, in the same topological arrangement, but often with large insertions. They have one active site carboxylic acid residue near the C-terminus of β strand 1, but lack the D or E found on strand 4, and the location of the other active site residues is quite variable.

The sequence homology among members of this family is extremely weak. When structurally aligned, the two closest members, MuA and HIV integrase, show only 15% sequence identity (Figure 6B). In fact, the homology is so poor that a 3D–1D profile search (Bowie et al., 1991), using a sequence profile based on the structure of MuA, failed to align the integrase sequence correctly, and vice versa.

Most of the few residues that are conserved between the two enzymes play conserved structural/functional roles. The role of two of these, N448 (N184 of integrase) and S463 (S195 of integrase), is shown in Figure 6C. The asparagine extends from helix E and its amide side chain forms hydrogen bonds to the amide backbone on the exposed edge of the β sheet, which may be important in the positioning of helix E relative to β strand 3. S463 is at the N-terminus of helix F, and its side chain forms both a direct hydrogen bond to the N-terminus of helix F and a water-mediated hydrogen bond to the C-terminus of helix E, thus bridging the tight corner between these two helices.

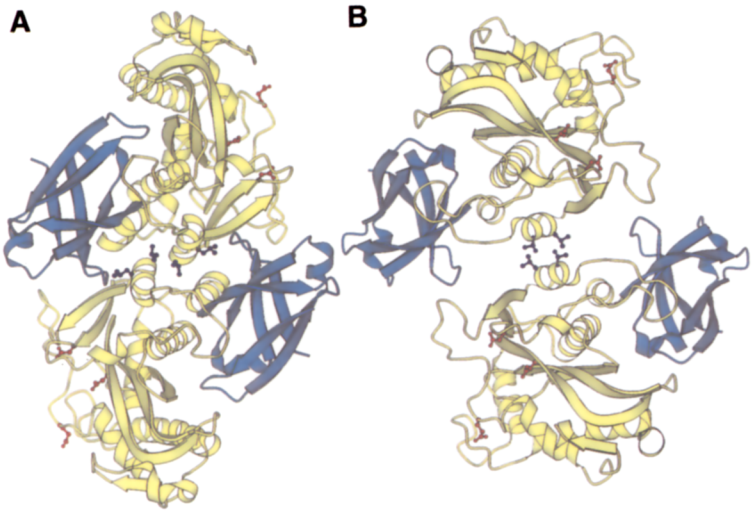

Figure 5. Dimer Interfaces

Subdomain 1 is colored yellow, subdomain 2 is blue, the active site residues are red, and L368 and L369, on helix C, are dark blue.

(A) The dimer formed by the two monomers in the asymmetric unit of the C2 crystal.

(B) A similar, but not identical, dimer formed across a crystallographic dyad axis in the $C222_1$ crystal. Although helix C is still near the center of the interface, details of the interaction are different in this crystal form, and less previously solvent-accessible surface area is buried.

Discussion

Location of the DNA

Although the sequence-specific DNA-binding domain has been removed, the core domain retains nonspecific DNA-binding activity (Nakayama et al., 1987). There are no canonical DNA-binding motifs in our structure. Electrostatic calculations show a large region of positive potential covering the lower side of the β barrel (Figure 7), suggesting that this region is involved in DNA binding despite the fact that it is not immediately adjacent to the active site. There are several possible segments of DNA with which the barrel could interact: the short stretch of DNA between the cleavage point and the endmost-specific MuA-binding site, the flanking host DNA, and the target DNA. The distance between the active site and the highly positive surface of the barrel suggests that it may act in *trans*, i.e., in the transpososome the barrel from one monomer may be adjacent to the active site of another.

Unfortunately, there is no detailed structural information on DNA complexes for any of the related polynucleotidyl transferases. Models of protein–nucleic acid complexes have been proposed for both Escherichia coli RNase H and RuvC (Katayanagi et al., 1992; Yang et al., 1990; Nakamura et al., 1991; Oda et al., 1993a; Ariyoshi et al., 1994). Although in detail the models are quite different, both place the substrate on the lower side of the enzyme, as drawn in Figure 6A. While a similar model appears reasonable also for MuA, these enzymes are sufficiently distinct from one another that it may not be valid to assume that the DNA approaches the active site in a similar manner in all cases.

Regulation of Catalytic Activity?

Although all three of the active site residues can be clearly seen in our electron density maps, they appear to be in an inactive configuration: the three carboxyl groups are too far apart to cooperatively bind to divalent metal(s). This is in contrast with the RNase H and RuvC structures in which the active site carboxyl groups are more tightly clustered. The position of E392 is the most striking: it is on a loop and pointed directly away from the other two residues. Although this loop is reasonably well ordered and is in nearly the same conformation in all three of our structures, there are several glycine residues in the vicinity and it is easy to imagine how it could be rearranged to bring the carboxyl of E392 into proximity with the rest of the active site. In the structure of the integrase core domain, this loop is too disordered to be modeled (Dyda et al., 1994), and in the structure of HIV RNase H, the corresponding loop is only ordered when it is part of the intact reverse transcriptase molecule (Davies et al., 1991; Jacobo-Molina et al., 1993; Kohlstaedt et al., 1992). Thus, flexibility in this loop region appears to be a common theme among some members of this group. Less obvious is that the distance between the two aspartic acid residues on the β sheet is greater in MuA than in integrase or in RNase H: the Cα atoms of D269 and D336 are 0.5 Å farther apart than the corresponding Cα atoms of RNase H (either E. coli or HIV) and 1.3 Å farther apart than the integrase ones. The side chain of D336 in MuA points away from D269 in our structures, further separating the two carboxyl groups.

As the protein does not appear to be active until it has formed a transpososome, there must be some mechanism to transfer information about DNA binding and the oligomerization state to the active site. Considering the distance of E392 from D269 and D336, one could imagine that each active site is constructed upon tetramerization from carboxylate residues contributed from separate MuA monomers. However, mixing experiments involving MuA with mutations at these acidic residues clearly speak against such a model (Baker et al., 1994). The structural similarities with other nucleotidyl transferases described here also disfavor a composite active site model and favor a model involving a cooperative structural transition upon tetramerization and substrate binding. From electrostatic considerations, we suspect that the β barrel is involved in binding to DNA, but it may also be involved in the regulation of the chemical activity of MuA. By comparison of the three independent structures of the core domain, we can

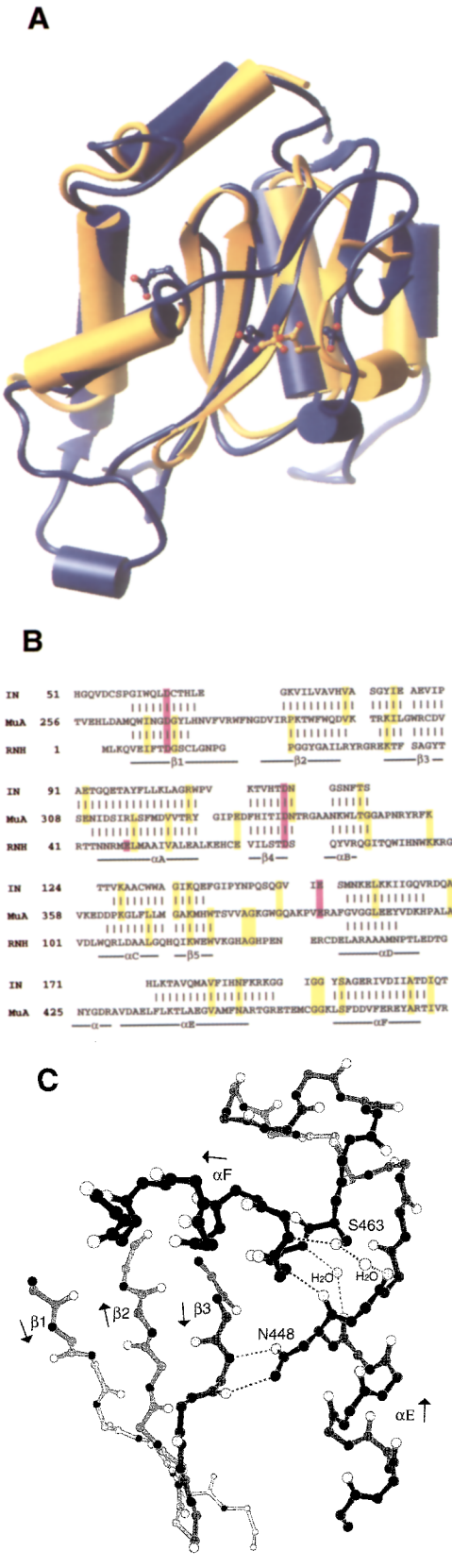

Figure 6. Comparison of MuA and HIV-1 Integrase

(A) Superposition of the MuA (blue) and HIV integrase (gold) cores. The MuA structure has been truncated after helix F. 114 Cα atoms superimpose on one another with an rms distance of 1.7 Å. Although all secondary structure elements found in integrase are also found in MuA, most of the turns in MuA are longer, some substantially.

(B) Structure-based sequence alignment. Sequences of the HIV-1 integrase core (IN), the MuA core, and E. coli RNase H (RNH). Vertical bars represent regions of corresponding structure. Helix D can only be approximately aligned. For RNase H, the alignment of all the helices is somewhat arbitrary. The three essential acidic residues in each enzyme are highlighted in pink. Other residues that are identical in both MuA and integrase or MuA and RNase H are highlighted in yellow. Secondary structure for MuA is noted underneath.

(C) Conserved architectural elements between MuA and HIV integrase. N448 and S463 are conserved and make similar, structurally important interactions, in both MuA and HIV integrase. Portions of β strands 1–3 and helices E and F are drawn as polyglycine, and the side chains of N448 and S463 are also shown. Oxygen atoms are shaded light gray; nitrogen is dark gray. Hydrogen bonds are shown by dotted lines. S463 bridges the tight corner between helices E and F, while N448 hydrogen bonds to the peptide backbone at the edge of strand 3 where it packs against the inside of helix E.

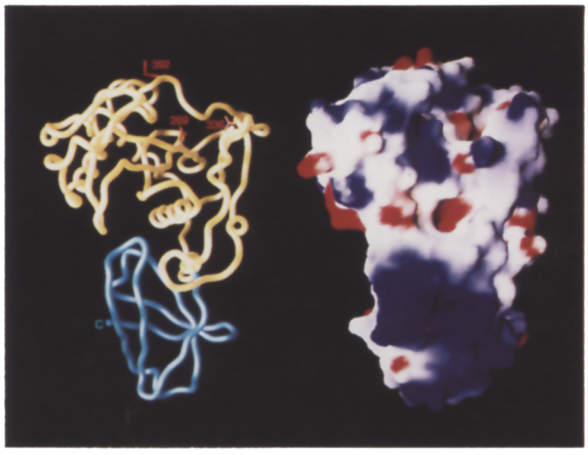

Figure 7. Electrostatic Potential

View is looking upward from the bottom of Figure 3B. Left, backbone trace, with subdomain 1 in yellow, 2 in blue, and the active site residues in red. Right, surface representation, from the same view, colored according to electrostatic potential, calculated by the program GRASP (Nicholls, 1993). Red represents −4.5 kT/e; white, neutral, and blue represent +4.5kT/e. A large region of positive potential is found on the surface of the β barrel, suggesting a role in DNA binding.

see that there is some flexibility in the relationship of the two subdomains. It could be that a larger change occurs on formation of the active complex (triggered by DNA binding to the barrel, by contacts with other subunits, or by both) and that the long loop between helices B and C transmits this back to the active site. In agreement with this hypothesis, the bond N-terminal of R355 appears to be a predominant chymotrypsin site in the tetramer compared with the monomer (T. Baker, personal communication). The peptide backbone of R355 is held in a nonideal geometry in our structure (Figure 2B). These observations point to a possibility that a significant structural change takes place at the loop that bridges the two subdomains upon formation of the MuA tetramer–DNA complex.

Interestingly, the HIV integrase core is active as an isolated domain in a disintegration assay: given a branched substrate mimicking the product of a strand transfer reac-

tion, it will reseal the target strand as the viral DNA segment is dissociated in an apparent reversal of the reaction (Bushman et al., 1993; Chow et al., 1992). We have failed to detect a similar activity with the MuA core (data not shown). It may be that in the integrase case, only substrate binding is necessary to bring the third residue into position, whereas MuA is more tightly regulated and requires higher order protein–protein interactions to bring this loop into the proper conformation.

It may be that the active site region does not even come into contact with the substrate until the transpososome has assembled: under conditions in which only simple binding occurs, the cleavage site at the Mu genome end and four to five base pairs inside of it are not protected from footprinting agents (Craigie et al., 1984; Zou et al., 1991). After formation of a stable transpososome, this region as well as approximately one helical turn of DNA outside of the Mu genome end are protected from footprinting agents (Lavoie et al., 1991; Mizuuchi et al., 1991). Thus, the required conformational change in the active site region may be coupled to a more global conformational change that brings the substrate into the active site.

Structural and Functional Similarities with Other Proteins

Although some similarity between the structures of MuA and HIV integrase was expected from the mechanistic similarities between the two, the degree of structural similarity seen is quite striking given the poor sequence homology. The functional relationship to RuvC and RNase H is less obvious, although both are nucleases that are sensitive to the local environment of the substrate: RuvC cleaves DNA strands involved in Holliday junctions, and RNase H cleaves only the RNA strand of RNA–DNA heteroduplexes.

The sequence homology between MuA and the other related proteins is so limited that it alone cannot provide a definitive answer as to how far beyond the first subdomain the structural homology will extend. The crystal structure of the HIV integrase core ends with helix F. The recently determined NMR structure of the C-terminal domain of the HIV integrase shows that, like the β barrel of the MuA core, it is rich in β strands, but their arrangement is distinct from that of MuA (P. Lodi, J. Ernst, M. Clore, and A. Groenenborn, personal communication). It is interesting to note that, as we believe for the MuA β barrel, the C-terminal domain of HIV integrase binds DNA nonspecifically (Vink et al., 1993; Woerner and Marcus-Sekura, 1993).

The contacts between the two subdomains of the MuA core cannot be well conserved among members of the transposase/integrase family, as a large fraction of them are provided by two loops of variable length. The loop between helices B and C extends from subdomain 1 to form a small but integral part of the barrel (Figure 4A). This loop is part of the spacer region between the second D and the E of the active site that is 35 residues in many members of transposase/integrase family, including HIV-1 integrase, but 55 in MuA. Members of the IS4 subfamily also appear to have longer spacer regions, but as these vary in length from 43–155 residues, details of this turn will vary widely among members of this subfamily (Rezsohazy et al., 1993). Another, smaller, contact surface between the two subdomains involves the loop between β strands 1 and 2, which is also too short in HIV integrase and many other related proteins to be a conserved contact.

MuA can also be activated by the MuB protein, which recognizes the final 36 residues of the MuA protein (Wu and Chaconas, 1994). The β barrel and the loop between helices B and C thus could also be involved in transferring information from the C-terminus of the protein (and MuB) back to the active site. However, several other mobile elements encode an ATPase protein analogous to MuB (Rowland et al., 1995; Radstrom et al., 1994). While it is not known if the transposases from these systems include a domain analogous to the β barrel of MuA, they have only a 35 residue spacer between the last two of the three conserved carboxylates and, thus, cannot have an extended loop bridging the two subdomains as shown in Figure 4A. MuA may have gained an additional mechanism to control its activity for the benefit of the lifestyle of Mu as a temperate phage in a way not critical for other elements. How the communication between the two subdomains is used for a physiological advantage may become clear when the structural organization of the transpososome complex becomes known.

More data is needed before we can justify modeling how four active sites are brought together in the transpososome to accomplish coupled cleavage and strand transfer. Although our structures do highlight two regions that are repeatedly used in protein–protein interactions, especially helix C, the exact nature of the interactions formed by these regions differs between the two different crystal forms. It is not known at this point whether these regions are responsible for interactions between monomers in the transpososome, or for interactions among different domains of the intact protein. Very little is known about the oligomerization state for other members of transposase/integrase family. Complementation studies indicate that HIV integrase is active as some form of oligomer (Engelman et al., 1994; van Gent et al., 1993), and its core domain is a dimer both in solution and in the crystal (Hickman et al., 1994; Dyda et al., 1994). However, the dimer formed by the integrase core cannot be replicated by the MuA core because helix G of MuA interferes with the positioning of helix E of the second monomer in the modeling of a similar dimer structure.

The sequence of MuA is a rather distant outlier in the transposase/integrase family. That it is so closely related structurally to HIV integrase implies that, at least in this domain, the structures of all members of the family are very similar. How far this homology will extend remains to be seen.

Experimental Procedures

Protein Purification, Crystal Growth, and Data Collection

MuA 248–574 was overexpressed in E. coli and purified essentially as previously described for the intact protein (Baker et al., 1993) with the following modifications: after $(NH_4)_2SO_4$ fractionation, a desalting column (Sephadex G-25) was used to equilibrate the redissolved precipitate with starting buffer for the phosphocellulose column. The KCl

Table 2. Statistics of the Data

	Resolution	R_{merge}[a]	Completeness	R_{cross}[b]
C2 form				
Native	2.7 Å	9.5%	92.5%	
Hg[c]	2.9 Å	12.3%	97.6%	10.2% (3.2 Å)
Pt[d]	3.5 Å	15.0%	92.2%	12.8% (3.5 Å)
$C222_1$ form				
Native	2.4 Å	10.8%	95.7%	

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$ in which I is an individual observation and $\langle I\rangle$ is the average of all observations of that reflection. For the two derivative data sets, all observations with $I > -3\sigma_I$ were used, for the C2 native, all those with $I > 0$, and for the $C222_1$ form, all those with $I > -3\sigma_I$.

[b] $R_{cross} = \Sigma|I_{native} - I_{derivative}|/\Sigma I$. All reflections with $I > 0$ were used in the calculation.

[c] Hg, 0.1 mM ethyl thio mercury salicylate.

[d] Pt, .01 × saturated platinum ethylene diamine dichloride.

gradient for the phosphocellulose column was from 0.4–1.2 M, and hydroxyapatite column chromatography was carried out without addition of KCl to the buffer. Purified protein was precipitated with $(NH_4)_2SO_4$ and redissolved at 16 mg/ml by dialysis against 10 mM Tris, 0.5 mM EDTA, 1 mM DTT (pH 7.5). Initial crystallization conditions were found using the commercial version (Hampton Research) of the sparse matrix screen (Jancarik and Kim, 1991). Crystals were grown by vapor equilibration using the hanging or sitting drop method. The C2 crystal form was grown as follows: well solution containing 13–16% (w/v) PEG 4000, 0.2 M Li_2SO_4, 75 mM Na_3 citrate (pH 6.0), and 2–15 mM DTT was mixed with protein (at 16 mg/ml, as above) in a 3:2 ratio and allowed to equilibrate at room temperature. The crystals were needles that often reach 2 mm in length, but were rarely thicker than 0.1 mm. The $C222_1$ form was grown from the same well solution, but with the addition of 0.5% (w/v) β-octyl glucoside and 5% PEG 400, and at 4°C rather than room temperature. These crystals grew as thin plates, and microseeding was necessary to obtain single ones. Both crystal forms were harvested into stabilizing solution of the same composition as the well solution except that the PEG 4000 concentration was raised to 20%.

Data was collected on an R_{axis} IIC image plate detector mounted on an RU200 rotating anode source at room temperature for the C2 crystals and at 4°C for the $C222_1$ crystals. A single crystal was used for each data set, although the C2 crystals were each translated two to four times during data collection. Data were reduced and scaled using DENZO and SCALEPACK (Otwinowski, 1993).

Structure Solution: C2 Crystal Form

For heavy atom soaks, crystals were briefly rinsed in a stabilizing solution containing no DTT before transfer to stabilizing solution containing heavy metal. The first derivative was prepared by soaking a crystal for 5 days in .01 × saturated (<1 mM) platinum ethylene diamine dichloride. The second derivative crystal was soaked overnight in 0.1 mM ethylthiomercury salicylate.

Phases were calculated and refined using MLPHARE (Otwinowski, 1991). Details of the data collection and phasing are shown in Tables 2 and 3. Initial phases were calculated to 3.2 Å and improved by solvent flattening using the routines within the CCP4 package (Wang, 1985; CCP4, 1979), by averaging about the noncrystallographic twofold and solvent flattening using the RAVE suite of programs (Kleywegt and Jones, 1994), and by averaging, solvent flattening, and histogram matching using SQUASH (Zhang, 1993). The location of the noncrystallographic twofold axis was originally defined using the PHASES package (Furey and Swaminathan, 1990) and later refined using the RAVE suite. Phases were extended to 2.9 Å using SQUASH or 3.075 Å using the RAVE suite. O was used for inspection of electron density and solvent masks and for model building (Jones et al., 1991). Several rounds of mask modification and averaging were necessary, as some turns were outside the original mask.

The model was refined by simulated annealing and energy minimization using XPLOR (Brünger, 1992). All data with $F > 1\ \sigma_F$ between 8 and 2.8 Å resolution were used in the refinement, except for a random 5% that was reserved for the free R factor calculation. The stereochemical parameters based on the Cambridge structural data base were used (Engh and Huber, 1991). Refinement started with one cycle of simulated annealing at 3000 K, followed by energy minimization and refinement of restrained individual B factors. No explicit noncrystallographic symmetry restraints were used, but during this first cycle the phases were restrained to those of the best averaged electron density map. This procedure reduced the R factor from 40.4% for the original model to 22% (29.7% for R_{free}). Subsequent rounds of rebuilding and energy minimization without phase restraints further reduced the R factor to 21.4% and 27.6% for R_{free}. The stereochemistry of the model is quite good, as can be seen in Table 1. No solvent is included in this model.

Structure Solution: $C222_1$ Crystal Form

This second crystal form was solved by molecular replacement using the refined coordinates of monomer 1 from the C2 crystal as a search model. XPLOR was used for both the rotation and translation functions. The rotation function, calculated between 15 and 3.5 Å resolution, gave a single peak 10.4 σ above the mean and 5.7 σ above the next highest peak. The translation function, calculated from 12 to 3.5 Å resolution, gave a peak 13.4 σ above the mean, also well above the next peak. The initial solution was refined as two rigid bodies, one corresponding to each subdomain, from 8 to 3 Å resolution, resulting in a drop in the R factor from 47.5% to 36.2%. After minor rebuilding, the resolution was extended to 2.5 Å and the model was subjected to simulated annealing at 3000 K. As before, 5% of the data was reserved for R_{free}. The model was then further improved by several

Table 3. Phasing Statistics

		Hg Derivative		Pt Derivative	
Resolution (Å)	Figure of Merit	Phasing Power	R_c	Phasing Power	R_c
10.27	0.4884	1.04	0.63	0.55	0.85
7.80	0.5024	1.09	0.60	0.84	0.74
6.30	0.5386	1.44	0.54	0.93	0.83
5.27	0.5251	1.54	0.54	0.85	0.99
3.98	0.4039	1.53	0.77	0.55	0.95
3.55	0.3236	1.43	0.86	0.51	0.98
3.20	0.2447	1.14	0.91	–	–
Total	0.3788	1.35	0.69	0.66	0.90

Heavy atom parameters were refined using MLPHARE (Otwinowski, 1991). For the Hg derivative (see Table 2), isomorphous difference data between 15 and 3.2 Å resolution and anomalous difference data between 15 and 5 Å were included in the calculations. For the Pt derivative (see Table 2), only isomorphous difference data was used, between 15 and 3.8 Å resolution. Phasing power is the ratio of the mean calculated heavy atom structure factor to the mean lack of closure error and is reported for the acentric reflections; R_c is the Cullis R factor, defined as the ratio of the mean lack of closure error to the mean isomorphous difference ($|F_{PH} - F_P|$), calculated for centric reflections only. Phases were later improved by solvent flattening and averaging about the noncrystallographic twofold (see text for details).

cycles of refinement and rebuilding, including one more cycle of simulated annealing, and the resolution was further extended to 2.4 Å. The turn from residue 418 to 425 was deleted, as there was very little electron density to support it, and 130 water molecules were added to the model. Statistics of the final model are presented in Table 1.

Miscellaneous Calculations and Figure Preparations

ACCESS was used to calculate solvent accessible surface areas (Lee and Richards, 1971). Figure 2 was made using the program O (Jones et al., 1991), Figures 3A and 6C were made using MAXIM, written and provided by Mark Rould, Figures 3B, 3C, 4, and 5 were made using MOLSCRIPT (Kraulis, 1991), Figure 6A was made using Ribbons (Carson, 1991), and Figure 7 was made using GRASP (Nicholls, 1993).

Acknowledgments

We thank Fred Dyda and David Davies for technical advice and Harri Savilahti for the clone of the MuA core. This work was in part supported by the National Institutes of Health intramural AIDS targetted antiviral program.

Received April 27, 1995; revised June 26, 1995.

References

Adzuma, K., and Mizuuchi, K. (1988). Target immunity of Mu transposition reflects a differential distribution of Mu B protein. Cell *53*, 257–266.

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a Holliday junction–specific endonuclease from E. coli. Cell *78*, 1063–1072.

Artymiuk, P. J., Grindley, H. M., Kumar, K., Rice, D. W., and Willett, P. (1993). Three-dimensional structural resemblance between the ribonuclease H and connection domains of HIV reverse transcriptase and the ATPase fold revealed using graph theoretical techniques. FEBS Lett. *324*, 15–21.

Baker, T. A., and Luo, L. (1994). Identification of residues in the Mu transposase essential for catalysis. Proc. Natl. Acad. Sci. USA *91*, 6654–6658.

Baker, T. A., and Mizuuchi, K. (1992). DNA-promoted assembly of the active tetramer of the Mu transposase. Genes Dev. *6*, 2221–2232.

Baker, T. A., Mizuuchi, M., and Mizuuchi, K. (1991). MuB protein allosterically activates strand transfer by the transposase of phage Mu. Cell *65*, 1003–1013.

Baker, T. A., Mizuuchi, M., Savilahti, H., and Mizuuchi, K. (1993). Division of labor among monomers within the Mu transposase tetramer. Cell *74*, 723–733.

Baker, T. A., Kremenstova, E., and Luo, L. (1994). Complete transposition requires four active monomers in the Mu transposase tetramer. Genes Dev. *8*, 2416–2428.

Betermier, M., Alazard, R., Lefrere, V., and Chandler, M. (1989). Functional domains of bacteriophage Mu transposase: properties of C-terminal deletions. Mol. Microbiol. *3*, 1159–1171.

Bowie, J. U., Luthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. Science *253*, 164–170.

Brünger, A. T. (1992). XPLOR Version 3.1: A System for X-Ray Crystallography and NMR (New Haven, Connecticut: Yale University Press).

Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993). Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc. Natl. Acad. Sci. USA *90*, 3428–3432.

Carson, M. (1991). Ribbons 2.0. J. Appl. Crystallogr. *24*, 958–961.

CCP4 (1979). The SERC (UK) Collaborative Computing Project Number 4, a Suite of Programs for Protein Crystallography (Warrington, England: Daresbury Laboratory).

Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992). Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science *255*, 723–726.

Clubb, R. T., Omichinski, J. G., Savilahti, H., Mizuuchi, K., Gronenborn, A., and Clore, G. M. (1994). A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase. Structure *2*, 1041–1048.

Craigie, R., and Mizuuchi, K. (1987). Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. Cell *51*, 493–501.

Craigie, R., Mizuuchi, M., and Mizuuchi, K. (1984). Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. Cell *39*, 387–394.

Davies, J. F. D., Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. Science *252*, 88–95.

Desmet, L., Faelen, M., Gama, M.-J., Ferhat, A., and Toussaint, A. (1989). Characterization of amber mutations in bacteriophage Mu transposase: a functional analysis of the protein. Mol. Microbiol. *3*, 1145–1158.

Doak, T. G., Doerder, F. P., Jahn, C. L., and Herrick, G. (1994). A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common D35E motif. Proc. Natl. Acad. Sci. USA *91*, 942–946.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science *266*, 1981–1986.

Engelman, A., Mizuuchi, K., and Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell *67*, 1211–1221.

Engelman, A., Hickman, A. B., and Craigie, R. (1994). The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J. Virol. *68*, 5911–5917

Engh, R., and Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. Acta Crystallogr. *A47*, 392–400.

Furey, W., and Swaminathan, S. (1990). PHASES: a program package for the processing and analysis of diffraction data for macromolecules. Am. Crystallogr. Assoc. Annu. Mtg. Program Abstr. *18*, 73.

Haniford, D. B., and Chaconas, G. (1992). Mechanistic aspects of DNA transposition. Curr. Opin. Genet. Dev. *2*, 698–704.

Haruki, M., Noguchi, E., Nakai, C., Liu, Y. Y., Oobatake, M., Itaya, M., and Kanaya, S. (1994). Investigating the role of conserved residue Asp134 in *Escherichia coli* ribonuclease H1 by site-directed random mutagenesis. Eur. J. Biochem. *220*, 623–631.

Hickman, A. B., Palmer, I., Engelman, A., Craigie, R., and Wingfield, P. (1994). Biophysical and enzymatic properties of the catalytic domain of HIV-1 integrase. J. Biol. Chem. *269*, 29279–29287.

Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P. et al. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl. Acad. Sci. USA *90*, 6320–6324.

Jancarik, J., and Kim, S. H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. J. Appl. Crystallogr. *24*, 409–411.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjelgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110.

Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., Inoue, H., Ohtsuka, E., and Ikehara, M. (1990). Identification of the amino acid residues involved in an active site of *Escherichia coli* ribonuclease H by site-directed mutagenesis. J. Biol. Chem. *265*, 4615–4621.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from *E. coli*. Nature *347*, 306–309.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1992). Structural details of ribonuclease H from *Escherichia coli* as refined

to an atomic resolution. J. Mol. Biol. *223*, 1029–1052.

Kim, K., Namgoong, S.-Y., Jayaram, M., and Harshey, R. M. (1995). Step-arrest mutants of phage Mu transposase. J. Biol. Chem. *270*, 1472–1479.

Kleywegt, G. J., and Jones, T. A. (1994). Halloween: masks and bones. In From First Map to Final Model, S. Bailey, R. Hubbard, and D. Waller, eds. (Warrington, England: Daresbury Laboratory), pp. 59–66.

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science *256*, 1783–1790.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Kruklitis, R., and Nakai, H. (1994). Participation of the bacteriophage Mu A protein and host factors in the initiation of Mu DNA synthesis *in vitro*. J. Biol. Chem. *269*, 16469–16477.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. *12*, 2331–2338.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lavoie, B. D., Chan, B. S., Allison, R. G., and Chaconas, G. (1991). Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu–host junction of the Mu type 1 transpososome. EMBO J. *10*, 3051–3059.

Lee, B., and Richards, F. M. (1971). The interpretation of protein structure: estimation of static accessibility. J. Mol. Biol. *55*, 379–400.

Leung, P. C., Teplow, D. B., and Harshey, R. M. (1989). Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. Nature *338*, 656–658.

Matthews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. (1967). Three-dimensional structure of tosyl-alpha-chymotrypsin. Nature *214*, 652–656

Mizuuchi, K. (1992). Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. *61*, 1011–1051.

Mizuuchi, K., and Adzuma, K. (1991). Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. Cell *66*, 129–140.

Mizuuchi, M., and Mizuuchi, K. (1989). Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. Cell *58*, 399–408.

Mizuuchi, M., Baker, T. A., and Mizuuchi, K. (1991). DNase protection analysis of the stable synaptic complexes involved in Mu transposition. Proc. Natl. Acad. Sci. USA *88*, 9031–9035.

Mizuuchi, M., Baker, T. A., and Mizuuchi, K. (1992). Assembly of the active form of the transposase–Mu DNA complex: a critical control point in Mu transposition. Cell *70*, 303–311.

Nakamura, H., Katayanagi, K., Morikawa, K., and Ikehara, M. (1991). Structural models of ribonuclease H domains in reverse transcriptases from retroviruses. Nucl. Acids Res. *19*, 1817–1823.

Nakayama, C., Teplow, D. B., and Harshey, R. M. (1987). Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. Proc. Natl. Acad. Sci. USA *84*, 1809–1813.

Namgoong, S. Y., Jayaram, M., Kim, K., and Harshey, R. M. (1994). DNA-protein cooperativity in the assembly and stabilization of Mu strand transfer complex. Relevance of DNA phasing and att site cleavage. J. Mol. Biol. *238*, 514–527.

Nicholls, A. J. (1993). GRASP Manual (New York: Columbia University).

Oda, Y., Iwai, S., Ohtsuka, E., Ishikawa, M., Ikehara, M., and Nakamura, H. (1993a). Binding of nucleic acids to *E. coli* RNase HI observed by NMR and CD spectroscopy. Nucl. Acids Res. *21*, 4690–4695.

Oda, Y., Yoshida, M., and Kanaya, S. (1993b). Role of histidine 124 in the catalytic function of ribonuclease HI from *Escherichia coli*. J. Biol. Chem. *268*, 88–92.

Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In Isomorphous Replacement and Anomalous Scattering, W. Wolf, P. R. Evans, and A. G. W. Leslie, eds. (Warrington, England: Science and Engineering Research Council, Daresbury Laboratory), pp. 80–86.

Otwinowski, Z. (1993). Oscillation data reduction program. In Data Collection and Processing, L. Sawyer, N. Isaacs, and S. Bailey, eds. (Warrington, England: Science and Engineering Research Council, Daresbury Laboratory), pp. 56–62.

Radstrom, P., Skold, O., Swedberg, G., Flensburg, J., Roy, P. H., and Sundstrom, L. (1994). Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. J. Bacteriol. *176*, 3257–3268.

Rezsohazy, R., Hallet, B., Delcour, J., and Mahillon, J. (1993). The IS4 family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. *9*, 1283–1295.

Rowland, S.-J., Sherratt, D. J., Stark, W. M., and Boocock, M. R. (1995). Tn552 transposase purification and *in vitro* activities. EMBO J. *14*, 196–205.

Surette, M. G., and Chaconas, G. (1991). Stimulation of the Mu DNA strand cleavage and intramolecular strand transfer reactions by the Mu B protein is independent of stable binding of the Mu B protein to DNA. J. Biol. Chem. *266*, 17306–17313.

Surette, M. G., and Chaconas, G. (1992). The Mu transpositional enhancer can function in trans: requirement of the enhancer for synapsis but not strand cleavage. Cell *68*, 1101–1108.

Surette, M. G., Buch, S. J., and Chaconas, G. (1987). Transpososomes: stable protein–DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. Cell *49*, 253–262.

Surette, M. G., Harkness, T., and Chaconas, G. (1991). Stimulation of the Mu A protein–mediated strand cleavage reaction by the Mu B protein, and the requirement of DNA nicking for stable type 1 transpososome formation: *in vitro* transposition characteristics of mini-Mu plasmids carrying terminal base pair mutations. J. Biol. Chem. *266*, 3118–3124.

van Gent, D. C., Vink, C., Groeneger, A. A., and Plasterk, R. H. (1993). Complementation between HIV integrase proteins mutated in different domains. EMBO J. *12* , 3261–3267.

Vink, C., Oude Groeneger, A. M., and Plasterk, R. H. (1993). Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. Nucl. Acids Res. *21*, 1419–1425.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. *115*, 90–112.

Woerner, A., and Marcus-Sekura, C. J. (1993). Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. Nucl. Acids Res. *21*, 3507–3511

Wu, Z., and Chaconas, G. (1994). Characterization of a region in phage Mu transposase that is involved in interaction with the Mu B protein. J. Biol. Chem. *269*, 28829–28833.

Yang, W. (1991). Structure determination of *Escherichia coli* selenomethionyl ribonuclease H by multiwavelength X-ray crystallography. PhD thesis, Columbia University, New York.

Yang, W., and Steitz, T. A. (1995). Recombining the structures of HIV integrase, RuvC and RNase H. Structure *3*, 131–134.

Yang, W., Hendrickson, W. A., Crouch, R. J., and Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. Science *249*, 1398–1405.

Zhang, K. Y. J. (1993). SQUASH: combining constraints for macromolecular phase refinement and extension. Acta Crystallogr. *D49*, 213–222.

Zou, A. H., Leung, P. C., and Harshey, R. M. (1991). Transposase contacts with Mu DNA ends. J. Biol. Chem. *266*, 20476–20482.

Protein Data Bank Accession Numbers

The accession numbers for the coordinates of the two crystal forms

reported in this paper are 1BCM and 1BCO and for the structure factors are R1BCMSF and R1BCOSF.

Note Added in Proof

The work cited as P. Lodi, J. Ernst, M. Clore, and A. Groenenborn, personal communication is now in press as Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1995). The solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry, in press. Also, the work cited as H. Savilahti, P. R., and K. M., unpublished data is now in press as Savilahti, H., Rice, P. A., and Mizuuchi, K. (1995). The phage Mu transpososome core: DNA requirements for assembly and function. EMBO J., in press.